Workgroup 5: PT Provider Assessment of Laboratory Performance

Mr. Daniel Tholen

Dan Tholen Statistical Consulting
Traverse City, Michigan, USA

Workshop Purpose

- Present issues where there are differences in different regions where PT is common
- Guide for development in other regions
- Deliberate redundancy and overlap of questions

- ✓ Discuss the issues
- ✓ Reach consensus or identify differences
- ✓ Report to the main group and journal

5-1-1a: WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF PT/EQA PROGRAMS THAT ARE MANAGED FOR EDUCATIONAL AND/OR REGULATORY PURPOSES?

- Threats of punishment cause a change in the way laboratories handle interlaboratory comparison samples.
- What resources are needed for educational activities?

5-1-1b: ARE THERE SOME WAYS TO SATISFY BOTH SETS OF NEEDS?

- Detection of poor performers
- Elimination of bad performers
- Elimination of ***bad performance***

- Information for lab self-improvement
- Information for method improvement

- **Please produce a short list of the relative features, and recommendations on how to "have it both ways".**

5-1-2a: WHAT ARE THE CONSIDERATIONS FOR DETERMINING WHEN TEST PERFORMANCE CAN BE GRADED?

- In USA, CLIA requires “all tests” to be included in PT and graded appropriately.
- “Regulated” vs. “Unregulated” analytes
- “Graded” vs. “ungraded” analytes
- Overall performance assessment

5-1-2b: SHOULD EVERY TEST BE GRADED, OR ARE THER SOME TESTS FOR WHICH PT/EQA IS PREMATURE, REDUNDANT, OR UNNECESSARY?

- What are the concerns for determining when an analyte is ready for grading, and what are the concerns for determining there is no need for PT?

Produce a list of concerns and analytes that could be excluded from PT.

5-1-2c: ARE THERE OBJECTIVE CRITERIA FOR MAKING THIS DECISION?

- For example, statistics such as
  - Interlab agreement, all results and by group.
  - Intermethod agreement
  - Number of laboratories participating
  - Likely proportion unacceptable

Or: 2+ PT organizers grading the analyte?

5-2-1a: WHAT PERFORMANCE MEASURES ARE APPROPRIATE FOR QUANTITATIVE AND QUALITATIVE TESTS?

- Accuracy (current)
- Short term precision (repeatability)
- Long term reproducibility
- Uncertainty
- Calibration or linearity
- Knowledge / interpretation
- Other?

5-2-1b: WHAT STATISTICAL TOOLS CAN BE USED TO MEASURE THESE CHARACTERISTICS?

- For measures other than accuracy, what would be appropriate statistical techniques?
  - Youden design
  - Repeat sample design
  - Linear design
  - ?

Please give a list, with preference order.

5-2-2a: WHAT CHARACTERISTICS OF PT/EQA PERFORMANCE CAN BE EVALUATED?

- This is supplemental to the previous question. Can PT be used to test skills such as interpretation or some aspects of handling?
- If so, how?

5-2-2b: IS IT POSSIBLE TO EVALUATE A LABORATORY'S INTERPRETATION OF TEST RESULTS?

- Provide a list of the best-supported ideas

5-3-1a: HOW SHOULD PERFORMANCE GOALS FOR LABORATORIES BE DETERMINED?

- Relative to others (SD, percentiles, rank)
- Expert Opinion – medical needs
- State of the Art
- Historical performance
- Other

5-3-1b: SHOULD PERFORMANCE BE MEASURED RELATIVE TO OTHER LABORATORIES, OR WITH OBJECTIVE GOALS?

- Relative to Others (common in EQA)
  - Z score
  - Percentiles
- Objective Goals (discouraged in CLIA)
  - Fixed limits
  - Percentage limits

5-3-2a: WHAT ARE THE CONSIDERATIONS FOR DETERMINING THAT A LABORATORY'S PERFORMANCE IS UNACCEPTABLE?

- Accreditation or regulatory requirements
- Use PT alone, or with what other data?

5-3-2b: SHOULD IT BE BASED ON A SINGLE TEST RESULT, A SET OF RESULTS IN A SINGLE TEST EVENT, OR RESULTS ACROSS SEVERAL TEST EVENTS?

- CLIA criteria:
  - 80% of samples for each analyte
  - 80% of samples for each subspecialty
  - 2 out of 3 consecutive PT events

NATA: Accuracy and Repeatability w/Youden design. No carryover.

5-3-2a: SHOULD PERFORMANCE GOALS BE THE SAME FOR ALL TYPES OF LABORATORIES?

- If PT goals are based on medical need, is it appropriate to have different criteria for different testing situations?
- How would performance needs be defined?

5-4-1: WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF "BLIND" PT/EQA,?

- Research findings
- Oversight concerns

- Designs where multiple pools are tested repeatedly

- Please report any shared group opinions

5-4-2: WHAT ARE THE CRITERIA THAT SHOULD BE USED TO DETERMINE THE FREQUENCY OF PT/EQA?

- Practical concerns (cost, turnaround time)
- Oversight concerns (undetected poor performance)
- Frequency of calibration
- Lab Workload

List criteria, preferences, objective measures

5-5-1: WHAT FACTORS SHOULD BE CONSIDERED IN DEFINING PEER GROUPS IF SUCH GROUPS ARE USED IN DETERMINING ACCEPTABLE LABORATORY PERFORMANCE?

- Tradeoff between group size and accommodating method bias
- Material matrix effects vs. calibration and method standardization (traceability)
- Need for verification?
- Consider education and regulation

5-5-2: WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF LABORATORIES RECEIVING IDENTICAL CHALLENGES IN EVERY TEST EVENT?

- Do all labs need to see the same test items at the same time?
- Advantages to alternate strategies?

- Consider alternative design strategies (i.e. random samples from pools, multiple lots and labs pick, blind, split sample, etc.)

Summary

- Purpose:
  - To discuss the issues
  - To reach consensus or identify differences
  - To report to the main group and journal
- Discuss other questions or tie-in to morning workshops.

SUBGROUP LEADERS AND RECORDERS

1. Dev Howerton
2. Tim O’Leary
3. Leigh Dini
4. Elizabeth Melnyk
5. Daniel Edson

- Suzette Park
- Mary Kimberly
- Darshan Singh
- Rex Astles
- John Hancock